REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

AD-A281 793

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
THESIS/~~DISSERTATION~~

4. TITLE AND SUBTITLE
A COMBINED BIASED-Robust Estimator for Dealing with Influence and Collinearity in Regression

5. FUNDING NUMBERS

6. AUTHOR(S)
James R. Simpson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Student Attending:
Arizona State University

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA-94-078

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STREET
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW 190-1
Distribution Unlimited
MICHAEL M. BRICKER, SMSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

29P 94-22701

94 7 19 174

14. SUBJECT TERMS

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT

18. SECURITY CLASSIFICATION OF THIS PAGE

19. SECURITY CLASSIFICATION OF ABSTRACT

20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

Thesis

Mark 'P' for allowance

Research Proposal

on

A Combined Biased-Robust Estimator for Dealing with Influence and Collinearity in Regression

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

James R. Simpson
Arizona State University

THE PROBLEM

Introduction and Background of the Problem

Regression analysis is a statistical tool that has earned widespread use in nearly all areas of endeavor seeking to fit a model to a set of data. Although there are several methods of estimating the model parameters, the least squares method is used most often because of its general acceptance, elegant statistical properties and ease of computation. Unfortunately, the mathematical elegance that makes least squares so popular depends on a number of fairly strong and many times unrealistic assumptions. The assumption that makes least squares so attractive in terms of hypothesis testing and confidence intervals on the parameter estimates is that the distribution of the errors is normal or Gaussian. This assumption can be violated if one or more sufficiently outlying observations are present in the data, resulting in less than optimal estimates of the parameters. The second problem that can ruin the accuracy of least squares estimates is correlated regressors. Highly correlated regressors can cause large variances in the estimates of the coefficients, sometimes resulting in incorrect levels of magnitude or even incorrect signs for the coefficients.

Outliers, which occur often in real data, occur for many reasons including typing or computation errors, interchanging of values, inadvertent observations from different populations and transient effects. Outliers can also be due to genuinely long-tailed distributions. Hampel et al. (1986) summarized the results of numerous studies of the frequency of outliers in real data and conclude that altogether 1-10% outliers in routine data are more the rule rather than the exception. Outliers can be found in the response variable (y-variable) or the regressor variables (x-variables). Regardless of the origin, a single, sufficiently outlying observation in a data set can render least squares estimation useless. Robust estimation methods can deal with outliers relatively easily. Ronchetti (1987) points out that the goal of a robust selection procedure is to choose a model which fits the majority of the data, taking into account that the errors may not be normally distributed. A number of robust regression estimation techniques have been proposed and some have been successfully used in practice.

Often when fitting a model to data, analysts find that some of the regressor variables are highly correlated with each other. This condition, known as multicollinearity, can have detrimental effects on the least squares estimates of the coefficients. In general, multicollinearity tends to inflate the

variance and absolute value of the least squares coefficients. In this case, the main problem with the least squares estimate is the restriction that the estimator be unbiased. Alternative estimation techniques that have been proposed successfully sacrifice small amounts of bias in exchange for large reductions in the variance of the estimates. Biased estimation methods, such as ridge regression, can provide stable coefficient estimates with computational ease.

Outliers and multicollinearity occur simultaneously in real data almost as often as each problem occurs separately. Relative to the amount of research in biased-only and robust-only techniques, the research in biased-robust regression has been sparse. Most of the advances in this area have been made in the last two decades by Holland (1973), Pariente and Welsch (1977), Hogg (1979), Askin and Montgomery (1980) Montgomery and Askin (1981), Pfaffenberger and Dielman (1984), Lawrence and Marsh (1984), Walker and Birch (1985, 1988), Walker (1987). Askin and Montgomery (1984) and Pfaffenberger and Dielman (1990) have followed up the development of their techniques by performing Monte Carlo simulation studies to compare various approaches. The most common approach to biased-robust estimation is augmented weighted least squares which allows a biased estimator and robust estimator to be combined into a single biased-robust estimator. Many of the existing robust estimators can be easily combined with biased estimators using the augmented-weighted least squares approach. In fact, several of the recently created biased-only and robust-only estimators are excellent candidates for an improved combined estimator.

Statement of the problem

Frequently, difficulties arise when practitioners try to apply appropriate regression estimation techniques. The traditional view that least squares is robust to deviations (even gross ones) from the assumptions of normality and uncorrelated regressors discourages users from applying other methods. In instances where the model adequacy diagnostics reveal a poor least squares fit due to outliers and collinearity, the practitioner is often not able to properly fit a model because the biased-robust estimation techniques are not known or available. The increasing presence of observational data with correlated regressors and abundant outliers makes advances in the state of the art of biased-robust estimation imperative. Although progress continues, there is a growing need for users to have tools available to implement when least squares fails. A need exists to develop and test alternative approaches to the combined problem so that the community of

practitioners are aware of the potential to accurately estimate regression model terms. The most recent advancements in robust-only and biased-only estimation warrant development of combined biased-robust estimators.

Research Objective

The objective of this research is to develop a biased-robust regression estimator and determine how the method performs in the presence of nonnormal errors (outliers) and multicollinear regressor variables. To accomplish this major objective a number of investigative questions must be answered. The sub-questions listed below are elements of the major objective and will guide the details of the research effort.

I. How will the biased-robust estimator be developed?

   A. What characteristics are required of the two classes of estimators (robust and biased) in order to take a robust estimator and a biased estimator and form a combined biased-robust estimator? Specifically, for each class of estimator:

      1. What are the strengths and weaknesses associated with the available techniques?

      2. What are the properties most desirable in an estimator?

      3. Which estimator is the best relative to the desirable properties?

   B. What characteristics are required for the biased-robust estimator?

      1. What are the properties most desirable for the combined estimator?

      2. Which estimator is the best relative to the desirable properties?

      3. What are the challenges associated with combining the estimators?

II. Which estimators should be used for comparison in the performance test?

III. How will each of these estimators be computed?

   A. Is software available that generates some of the chosen estimators?

   B. Which estimators require coding?

   C. Which programming language is most appropriate to code the remaining estimators?

IV. How will the Monte Carlo simulation be developed to compare the estimators?

   A. What characteristics of the data are important to vary in the simulation?

   B. What type of design will be used in this experiment?

   C. How will the data be generated?

V. What criteria will be used to measure the performance of the biased-robust estimators?

   A. What performance indices are important?

   B. What measures can be calculated based the simulation results?

Scope and limitations of the study

- A subset of the most promising robust and biased estimation techniques will be modeled and compared.
- Monte Carlo simulation will be used to compare the techniques. A designed experiment will be developed to test the estimation technique in the presence of a number of different types of data.
- The primary purpose of this study is **not** to **only** identify an estimator with the superior statistical properties. Certain statistical properties such as high breakdown point are important and will be treated accordingly. In addition, the estimators that have some asymptotic distributional properties is preferred because parametric tests of hypothesis can be performed. Of equal importance though are the method's ability to accurately estimate the model coefficients. Overall assessments will be based on the combined knowledge of statistical properties **and** performance results against data from the experiment.

Outline of the remainder of the paper

- Review various robust estimators, biased estimation techniques and biased-robust estimators
- Methodology detailing the proposed combined estimator and its properties
- Determination of computational procedures for the estimators, design of the experiment, generation of the data, and identification of the measures of performance used in the Monte Carlo simulation

REVIEW OF THE RELATED LITERATURE

In general, the majority of the research on alternatives to least squares estimation in the presence of outliers and correlated regressors has addressed either the nonnormal issue or the collinearity issue but seldom addressed the combined problem. This review will cover the three topics in proportion similar to the amount of research available in the literature. There are two reasons: 1) in this case it is true that the more research that has been performed, the more significant are the findings, 2) a thorough understanding of the biased-robust estimation problem is aided by one becoming familiar with the work in robust-only and biased-only estimation. The contributions to biased-robust estimation follow naturally and will be discussed in detail concerning both the estimation approaches and the Monte Carlo simulation comparisons.

Robust Estimation

The problem of robustness in statistics goes back to the beginnings of statistics, especially in terms of measures of location. In fact, Rey (1983) notes that the Greek besiegers of antiquity switched from using the mean to a more robust measure, the median. Hampel et al. (1986) point out that rejection of outliers was considered by Bernoulli (1777) and Bessel and Baeyer (1838). Formal rejection rules were given by Peirce (1852) and Chauvenet (1863). Thorough accounts of the early

work can be found in papers by Harter (1974-1976), Huber (1972), and Stigler (1973). It was not until recent decades though that robust estimation became a true research area. The awareness was created by people such as E. S. Pearson, G. E. P. Box and J. W. Tukey. Box (1953) actually coined the term *robustness* and Tukey (1960) demonstrated the drastic nonrobustness of the mean and presented robust alternatives. In the 1960s, papers by Huber (1964, 1965, 1968) and Hampel (1968) formed the basis for the theory of robust estimation and extended this theory to applications such as regression.

Since these pioneering papers on robust estimation in regression, many approaches have been presented but no single approach is either optimum or superior to the others in all aspects. The important criteria used in the field to determine the strengths and weaknesses of an estimator will be introduced prior to the discussion of each of the techniques. Although some of the criteria are more important than others for a particular set of data, the optimum estimator would ideally have the positive characteristics of all criteria.

Equivariance: Refers to statistics that transform properly. It can be one of three types: affine, scale or regression equivariant (Rousseeuw and Leroy, p. 116). Affine equivariance means that, under the sum of a linear transformation and a fixed vector, the estimator is transformed in the same way. Scale equivariance means that if the observations are multiplied by a constant c, the estimators are also multiplied by c. Regression equivariance means that without loss of generality, $\beta=0$.

High breakdown point: The breakdown point of an estimator is the amount of contamination allowed in the data (usually a percentage or fraction) until the estimate ceases to give information about the parameters. Breakdown points can be as low as 0% (or sometimes referred to as 1/n) meaning that only a single outlying observation can cause an estimator to be meaningless, as is the case with least squares. Breakdown points can be as high as 50%, meaning that up to half of the data can be contaminated and the estimator can still be useful.

Efficiency: Expressed as a percentage, the degree to which the estimator performs like least squares in the presence of Gaussian or normally distributed errors. The term is computed as the mean squared error of the robust fit divided by the mean squared error of the least squares fit. Efficiencies near 90-95% are desirable.

X-space outlier: An unusual point in the x-direction. Its effect on a least squares estimator is very large because it "pulls" the least squares line in its direction. For this reason this observation is also called a leverage point.

Y-space outlier: An unusual point in the y-direction only. This point can have a large influence on the least squares line but the nature and extent of the effect depends on its x-coordinates and the disposition of the other points. It is important to note that the most dangerous type of point is one that is an outlier in both directions (*x and y*-space outlier).

Computational ease: Considerations include the complexity and availability of the method used to calculate the estimates. This measure also considers the potential for convergence problems.

Distributional properties: In order to test the adequacy of the estimation technique and choose the parameters which are significant in the model, hypothesis tests must be performed. These tests are more efficient if they are based on some, at least asymptotic, assumptions about the distribution of the estimator.

A graphic will be displayed next to each robust technique discussed that quickly highlights the strengths and weaknesses of the method using the criteria just mentioned. Strengths will be indicated by shading.

L_1-norm or (least absolute values) estimation

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

Many alternative estimators have been proposed for regression. One of these approaches came from Edgeworth (1887), improving a proposal of Boscovich (1757). He proposed the L_1-norm or least absolute values (LAV) regression estimator, which is determined by

$$\min \sum_{i=1}^{n} |r_i| \tag{1}$$

This approach attempts to minimize the sum of the absolute errors. The LAV estimator is commonly solved with linear programming methods. Unfortunately, the breakdown point of LAV regression is still no better than 0%. The LAV is robust to an outlier in the y-direction (unlike least squares). However, LAV regression does not protect against outlying x, where the effect of the leverage point is even stronger than on the least squares line. It turns out that when the leverage

point is far enough away, the LAV line passes right through it. So a single erroneous point can totally offset the LAV estimator.

The L_1-norm and least squares (L_2-norm) are special cases of the L_p-norm regression problem. The objective in the general case is to

$$\min \sum_{i=1}^{n} |r_i|^p \tag{2}$$

where $1 \le p \le 2$. This approach has been considered by Gentlemen (1965), Forsythe (1972) and Sposito et al. (1977). Dodge (1984) suggested a regression estimator based on the convex combination of the L_1 and L_2 norms. All these proposals possess a zero breakdown point.

M-estimation

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

Huber (1973) introduced a class of estimators called "M-estimators". This method is the most popular of all robust estimators. The M-estimators are based on the idea of replacing the squared residuals by another function of the residuals ρ(r), where ρ is a symmetric function with a unique minimum at zero.

$$\min_{\beta} \sum_{i=1}^{n} \rho(e_i) = \min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i'\beta) \tag{3}$$

M-estimators are maximum likelihood estimators where the function ρ is related to the likelihood function for an appropriate choice of the error distribution. Because the M-estimator is not scale invariant the minimization problem is modified by dividing the ρ function by a robust estimate of scale s, so the formula becomes

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i'\beta}{s}\right) \tag{4}$$

A popular choice for s is

$$s = median\ |e_i - median(e_i)|\ /\ 0.6745$$

The constant 0.6745 is used to make s an unbiased estimator of σ when n is large and the sample actually arises from a normal distribution.

The least squares estimator is a special case of the $\rho(\ )$ function where $\rho(u) = \frac{1}{2}u^2$. For a convex ρ, equivalence to (4) can be found by finding the first partial derivatives of (4) with respect to β and setting the result equal to **0**, as

$$\min_{\beta} \sum_{i=1}^{n} \psi\left(\frac{e_i}{s}\right) = \min_{\beta} \sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{s}\right)\mathbf{x}_i = \mathbf{0} \tag{5}$$

where $\psi(u) = \frac{\partial}{\partial u}\rho(u)$, resulting in the necessary condition normal equations. If $\psi(u)=u$, then (5) reduces to the normal equations yielding the least squares estimator. However, in the case of robust estimation, $\psi(u)$ is not linear so that (5) defines a nonlinear system of equations which requires an appropriate iterative technique.

The $\psi(u)$ function controls the weight given to each residual and is very important in determining the robust and efficiency properties of the estimator. Although a number of popular ψ-functions have been developed, they primarily belong to one of two categories: monotonic and redescending. The least squares ψ-function described earlier reveals its weakness in situations involving heavy-tailed distributions. The ψ-function, $\psi(u)=u$, is unbounded meaning large residuals receive heavy weights. The Huber function (Huber, 1964), is an example of a monotone ψ-function defined as $\psi(u) = min(c_H, max(u, -c_H))$ which results in down-weighting the large residuals compared to least squares. Other ψ-functions redescend with increasing residual magnitude. The bisquare or biweight function of Beaton and Tukey (1974), is defined as $\psi(u) = u(1-(u/c_B)^2)^2$ for $|u|<c$ and 0 if $|u| > c$. The c terms in both equations refer to tuning constants chosen to achieve desired efficiencies. The values c_H=1.345 and c_B=4.685 for the Huber and biweight ψ-functions respectively achieve 95% efficiency compared to the least squares estimator in the model when the errors are actually normally distributed. For an excellent summary of different approaches to the ψ-functions, see Montgomery and Peck (1992).

The solution to (5) requires solving a system of equations using iteration schemes. Approaches include reweighted least squares, or the so-called H-algorithm. Iteratively reweighted least squares

(IRLS) is the most widely used nonlinear optimization technique in robust regression. Based on a starting value β_0, the iteration scheme is found by

$$\beta_1 = \beta_0 + (X^T < w(\frac{y - X\beta_0}{s})X)^{-1} X^T < w(\frac{y - X\beta_0}{s}) > (y - X\beta_0) \tag{6}$$

A major reason for the widespread application of IRLS is that it can be used in an ordinary or weighted least-squares framework. This can be demonstrated by expressing the above form as

$$\mathbf{X'WX}\beta = \mathbf{X'W}y \tag{7}$$

where W is an n x n diagonal matrix of weights with diagonal elements $w_1, w_2, \ldots, w_n$ given by

$$w_i = \frac{\psi[(y_i - x_i'\beta_0)/s]}{(y_i - x_i'\beta_0)/s} \tag{8}$$

The equation in (7) results in the usual weighted least squares normal equations. Thus, the one-step M-estimator can be found at convergence, where

$$\beta_1 = (X'WX)^{-1} X'Wy \tag{9}$$

At each iteration the weights are recomputed using the updated estimate of β. After the first iteration $\hat{\beta}_0$ is replaced by the updated estimate β_1. Usually only a few iterations are required to achieve convergence.

One may be interested in the distributional properties of β. Huber (1981) showed that, under certain conditions, the asymptotic distribution of β is $N(\beta, V_A)$, where V_A is a function of σ^2, the ψ-function and its derivative. Unfortunately, the finite sample distribution of β and its covariance matrix is not known. Holland and Welsch (1977) point out that one approach to robust inferential procedures based on β utilizes finite sample approximations to V_A. They discuss several alternative finite sample estimates of the covariance matrix of β.

Concentrations in research have focused on the best technique for solving the system of equations. IRLS is the most popular approach, but subtleties in the approach are still unresolved. In each step of the iteration procedure, both the coefficients and the scale can be simultaneously

reestimated. Convergence concerns arise when the scale estimate is reestimated. Some authors suggest iterating on scale (Rousseeuw and Leroy 1987; Street et al. 1988), while others suggest fixing the scale estimate (Hogg 1979; Green 1984). It is also very important to start the iteration with a "good" starting value, one that is already sufficiently robust. Without this precaution one can easily end up in a local minimum that does not correspond at all to the expected robust solution. The calculation of bounded influence estimators presents similar problems.

M-estimators have taken the art of robust estimation to a higher, more applicable level. Vast amounts of research has been conducted constructing the ψ-functions so that the estimators are both robust and efficient. M-estimators are statistically more efficient than LAV regression, while at the same time they are robust with respect to outlying y. However, as will be discussed later in the section on bounded-influence methods, M-estimators are not robust to x-outliers. Also, their breakdown point is 1/n because of the effect of a single outlying x.

R-estimation and L-estimation

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

R-estimates are based on the ranks of the residuals. The idea of using these in multiple regression is attributed to Adichie (1967), Jaeckel (1972), and Jureckova (1977). The proposal of Jaeckel uses the rank R_i of the residual $r_i = y_i - x_i\beta$ in the objective function as

$$\min \sum_{i=1}^{n} a(R_i) r_i \tag{10}$$

where a(i) is the scores function. Examples of scores functions are the Wilcoxon scores and median scores.

Several research efforts have focused on using a linear combination of order statistics to obtain a robust estimate called an L-estimator. The order statistics of a random sample of a continuous distribution are $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$, where $x_{(i)}$ is the i^{th} order statistic. Bickel (1973) has proposed a class of one-step estimators for regression that depend on an initial estimate of β. Koenker and Bassett (1978) use analogs of sample quantiles for regression. The trimmed least squares of Ruppert and Carrol (1980) are also L-estimators.

The performance of R- and L- estimators have not been as good as the M-estimators for the regression problem (Heiler, 1981).. Montgomery and Peck point out that L-estimators do not always generalize clearly to multiple regression and both R- and L-estimates are more computationally difficult to obtain than M-estimates.

Least Median of Squares (LMS) estimation

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

Instead of using least sum of squares, which can also be interpreted as least squares on the *mean*, what about least squares on the *median*? This approach was first proposed by Hampel (1975, p. 380) and was later adopted and refined by Rousseeuw (1984). Rousseeuw proposed the least median of squares (LMS) estimator given by

$$\min_{\beta} med\ r_i^2 \tag{11}$$

This estimator is robust with respect to outliers in both the x- and y-directions. Its breakdown point is the highest possible (50%) and the estimator is equivariant. Unfortunately, the LMS estimator is not efficient relative to least squares when the errors are normal. Also, the computational effort involves evaluating all possible 2-point subsets and using the estimate that produces the small median squared residual. This approach can result in the estimate being adversely effected by outliers Because of its low efficiency, Rousseeuw and Leroy suggest using it for data analytic purposes (detecting outliers) or as an initial stage estimator.

Least Trimmed Squares (LTS) estimation

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

The least trimmed squares (LTS) approach was developed also by Rousseeuw (1983) as a high efficiency alternative to LMS. The LTS estimator is given by

$$\min_{\beta} \sum_{i=1}^{h} (r^2)_{i:n} \tag{12}$$

where $(r^2)_{1:n} \leq (r^2)_{2:n} \leq \ldots \leq (r^2)_{n:n}$ are the ordered squared residuals and h is the number of residuals included in the calculation. This approach is similar to least squares except the largest α squared residuals are not used (trimmed sum) in the summation, allowing the fit to avoid the

outliers. This approach converges at a rate similar to the M-estimators. It is also equivariant and the breakdown point is 50% when h=n/2. According to Rousseeuw and Leroy, the main disadvantage of LTS is the large number of operations required to sort the squared residuals in the objective function. Another challenge is deciding the best approach for determining the initial estimate.

Bounded-influence or generalized M-estimators

Equivariance
High Breakdown Point
Efficiency
X-space Outliers
Y-space Outliers
Computational Ease
Distributional Properties

The M-estimator can successfully handle situations where the outliers in the response variable occur at points in the regressor space with low to moderate leverage. Outliers occurring outside the regressor space in either the response variable or independent variable direction at high leverage locations create problems not only for the least squares estimator, but for the M-estimator as well. In particular, M-estimators are vulnerable to points having a small residual with the corresponding leverage or influence on the regression equation being very large. These small residual, high leverage points could receive full weight under M-estimation.

The diagonal elements of the "hat matrix" $H=X(X'X)^{-1}X'$, denoted h_{ii} are typically used as measures of leverage. The h_{ii} is a standardized measure of the distance of a point x_i' to the centroid of the regressor space. The range of h_{ii} is $1/n \le h_{ii} \le 1$ and the average value of h_{ii} is p/n. Hoaglin and Welsch (1978) suggest that values of $h_{ii} > 2p/n$ can be considered high leverage points.

A robust technique that attempts to downweight the high influence points as well as large residual points is bounded-influence (BI) estimation. The BI estimators are solutions to the normal equations formed from

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - x_i'\beta}{s\pi_i}\right) x_i = 0 \qquad (13)$$

where, for appropriate values of π_i the BI estimator can downweight outliers with high leverage points. The estimator described here was developed by Schweppe (see Hill, 1977). The other main type of BI estimator was proposed by Mallows (1975). The distinction between these two types is that the Mallows estimator does not have the π weight in the denominator of the ψ-

function. Both types have the effect of downweighting leverage points, but the Schweppe weighting scheme downweights only if the residuals are large. Krasker and Welsch (1982) describe a weakness in the Mallows estimator:

> Outlying points in the X space increase the efficiency of most estimation procedures. Any downweighting in X space that does not include some consideration for how the y values at the outlying observations fit the pattern set by the bulk of the data cannot be efficient.

They go on to say that the Schweppe estimator has the potential to overcome these efficiency problems.

IRLS can be used again to solve (13). At convergence, the BI estimator can be written as

$$\beta_{\text{BI}} = (X'WX)^{-1}X'Wy \tag{14}$$

where in this case the diagonal elements of W are the weights w_i defined as

$$w_i = \frac{\psi[(y_i - x_i'\beta_{BI})/\pi_i s]}{(y_i - x_i'\beta_{BI})/\pi_i s} \tag{15}$$

Several authors, including Krasker and Welsch (1982) suggest that the π_i take the form

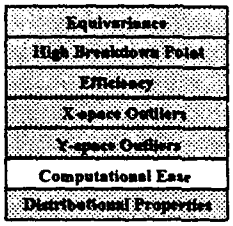
$$\pi_i = [(1 - h_{ii})/h_{ii}]^{1/2} \tag{16}$$

Several suggestions for the π-weights have been made that involve typical least squares outlier diagnostics including DFFITS used in (16) above. Other suggestions include studentized residuals, PRESS residuals or even Cook's D statistic. Each of these diagnostics measures leverage to some degree because each contains h_{ii} in their respective equation. Suggestions for the ψ-functions include various different M-estimate approaches such as Huber's t and Tukey's biweight. The research in this area is fairly new and some of the untried combinations of π-weights and ψ-functions could produce excellent estimators.

Bounded-influence estimators posses the same efficiency and asymptotic distributional properties of M-estimators. The breakdown point of the BI approach improves on the 1/n value of M-estimation, but is still not considered a high breakdown point estimator. The breakdown point is a function of the number of variables p, and is no greater than $1/p$. This condition can lead to

problems in models with many regressors. Also, both M-estimation and BI estimation can be improved by starting with a good initial estimate.

Multi-stage robust estimators



The discussion of robust estimators has clearly shown that no estimator has all of the desirable properties. Some of the methods have been proposed to obtain good initial estimators, while others reveal that they can be enhanced by a good initial estimate. The idea behind multi-stage robust estimators is to take advantage of these complementary needs. The purpose is to use different techniques in different stages so that the desirable properties of each technique can be combined. For example, if an LMS estimator can be effectively combined with a BI estimator and the properties maintained, then an estimate could be developed that is equivariant, efficient, has a high breakdown point, bounds the influence and has asymptotic distributional properties. Although this idea has been around for a few years (Hampel et al. 1986; Rousseeuw and Leroy 1987; and Ronchetti 1987), only in the last year or so have techniques actually been developed. Simpson et al. (1992) and Coakly and Hettmansberger (1993) propose two stage estimators that use high breakdown point estimators to generate a starting value and bounded-influence estimation to find the final value. Simpson et al. used an LMS initial stage and bounded-influence (Mallows type) second stage estimator. Coakly and Hettmansberger propose an LTS initial estimate, followed by a Schweppe type bounded-influence estimator.

Both approaches use a one-step estimation method to solve the system of equations for the second stage estimate after finding the initial estimate $\hat{\beta}_0$. They both use a one-step Newton-Raphson method. Simpson et al. state that one-step estimation inherits the breakdown point of the initial estimator and at the same time maintains the sample distribution of the *secondary* estimate. They say that IRLS inherits the asymptotic distribution of the *initial* estimate. More investigation is required here to determine the best approach to use in solving for the second stage estimate.

Coakly and Hettmansberger show that their estimator satisfies the goals of high breakdown, bounded-influence, and high efficiency. They also derive the asymptotic sampling distributions showing that the estimator is asymptotically normal, similar to the fully iterated general M-estimator.

This multi-stage approach to robust estimation clearly shows the most promise. Many different choices of estimators are available for each of the stages. The methodology discussed in the following chapter will describe some of the possibilities.

Biased estimation

The review of the literature in this section will not be nearly as detailed as the robust topics because the techniques in biased-estimation are fairly well-known and proven to be q successful. Some recent research describing slight modifications to the approaches will also mentioned. The techniques associated with biased estimation that have been used by those modeling the combined influence-collinearity problem have mostly involved ridge or generalized ridge regression. Askin and Montgomery (1984) showed that some of the other techniques, including principal components regression and Stein shrinkage, were consistently outperformed by ridge and generalized ridge. These two techniques will be briefly described. Most of this introductory information is contained in Montgomery and Peck, who present a complete summary of the approaches to biased estimation.

Ridge Regression

Ridge regression is the most popular and commonly used method for dealing with multicollinearity. The objective is to reduce the size and variance of the least squares estimates by introducing a slight amount of bias. This approach was originally proposed by Hoerl and Kennard (1970a, b). The ridge estimator is determined by solving a modified version of the least squares normal equations. The ridge estimator, $\hat{\beta}_R$, is given by

$$\hat{\beta}_R = [\mathbf{X}'\mathbf{X} + k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{y} \tag{17}$$

where $k \geq 0$ is called the biasing parameter and is selected by the analyst. The challenge in this approach is finding the appropriate selection of k. Many methods for choosing k have been proposed. The approach recommended by Hoerl and Kennard (1970a) is to choose k by inspection of the ridge trace. The objective is to select the smallest value of k in which the estimate of $\hat{\beta}_R$ stabilizes. Other suggested approaches that are more analytical have been proposed by Hoerl and Kennard (1976), McDonald and Galarneau (1975), and Mallows (1973). If the analyst's primary

purpose in developing a model is prediction, Montgomery and Friedman (1993) propose choosing k that minimizes $\text{PRESS}_R(k)$ which is the PRESS statistic calculated for the ridge estimator.

An import computational aspect of ridge regression is that the estimates may be found by using an ordinary least squares program and augmenting the standardized data. This approach gives

$$X_A = \begin{bmatrix} X \\ \sqrt{k}I_p \end{bmatrix} \qquad y_A = \begin{bmatrix} y \\ 0_p \end{bmatrix} \tag{18}$$

where $\sqrt{k}I_p$ is a p x p diagonal matrix with diagonal elements equal to $\sqrt{k}$. The associated ridge estimates are computed by

$$\hat{\beta}_R = [\mathbf{X}'_A\mathbf{X}_A + k\mathbf{I}]^{-1}\mathbf{X}'_A\mathbf{y}_A = [\mathbf{X}'\mathbf{X} + k\mathbf{I}_p]^{-1}\mathbf{X}'\mathbf{y} \tag{19}$$

This augmented matrix approach can be used effectively with iteratively reweighted least squares to form a combined biased-robust estimator.

Generalized Ridge Regression

Generalized ridge regression is an extension to ridge that was proposed by Hoerl and Kennard (1970a) that allows separate biasing parameters to be obtained for each regressor. Working with a model transformed to the space of orthogonal regressors simplifies the discussion. Assuming **T** is the orthogonal matrix of the eigenvectors of the **X** matrix, let **Z**=**XT** and α=**T**'β so that α become the transformed model coefficients. The generalized ridge coefficients become

$$\hat{\beta}_{GR} = \mathbf{T}\hat{\alpha}_{GR} \tag{20}$$

The mean square error is minimized by selecting $k_j = \sigma^2 / \alpha_j^2$. Several authors, including Hemmerle (1975) noted that choosing k_j in this fashion results in too much shrinkage. Montgomery and Peck suggest constraining the maximum increase in the residual sum of squares to between 1 and 20 percent. This approach was used by Askin and Montgomery (1984) in their analysis of augmented robust regression procedures.

Biased-robust Estimation

As was mentioned previously, the study of estimation under the simultaneous problems of influence points and collinearity has not been researched nearly in as much depth as either of the single problems. In fact, it wasn't until 1973 that Holland introduced the first approach to estimating under the simultaneous conditions. Later, Askin and Montgomery (1980) introduced a family of estimators that combined robust M-estimation criteria with biased estimation constraints. Pfaffenberger and Dielman (1984) used a similar approach but replaced the M-estimate with LAV estimation. Lawrence and Marsh (1984), Askin and Montgomery (1984), and Pfaffenberger and Dielman (1990) compare alternative combinations of ridge regression and robust regression techniques. Askin and Montgomery, and Pfaffenberger and Dielman use designed experiments with Monte Carlo simulation, while Lawrence and Marsh use real data to predict fatalities in the US coal mining industry. Walker (1987) modified Askin and Montgomery's approach to allow bounded-influence estimators to be used instead of M-estimators, thus being able to better control the influence. Walker emphasizes the importance of applying these types of estimators in the combined problem by showing the potential effects of collinearity on robust estimators and also the effects of influence on biased estimators.

The approach suggested by Hogg (1979) and Askin and Montgomery (1980) was to apply some sort of robust estimation to a ridge regression model. The ridge estimator is first obtained by augmenting the least squares design and observation matrices with p additional rows. The robust ridge estimators are the solution to the problem

$$\min_{\beta} \sum_{i=1}^{n} \rho(y_i - \mathbf{x}_i'\tilde{\beta}) \tag{21}$$

$$\text{subject to } \tilde{\beta}'\tilde{\beta} \le d^2$$

where the objective function is the classic M-estimator described previously. The solution to this problem can be obtained by IRLS where the weights on augmented observations are fixed at 1.0. The resulting estimator becomes

$$\tilde{\beta} = [\mathbf{X}'\mathbf{W}^k\mathbf{X} + k\mathbf{I}]^{-1}\mathbf{X}'\mathbf{W}^k\mathbf{y} \tag{22}$$

where ${}_k = diag(w_1^k, w_2^k, \ldots, w_i^k)$ and $w_i^k = \psi(e_i^k / s)/(e_i^k / s)$. The weighting matrix is now a function of the shrinkage parameter k. Sensitivity to influential observations is a problem because M-estimators are used.

A natural extension of augmented robust estimators are augmented bounded-influence estimators (Walker, 1987). The estimator in this case is the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \mathbf{x}_i'\tilde{\beta}}{\pi_i s}\right)\pi_i \tag{23}$$

$$\text{subject to } \tilde{\beta}'\tilde{\beta} \le d^2$$

where the objective function is now the bounded-influence estimation approach. The estimator is found using (22) above, but in this case the weights are found by applying the bounded-influence approach where $w_i^k = \psi(e_i^k / \pi_i s)/(e_i^k / \pi_i s)$. The weights in this case are not fixed, but are functions of the shrinkage parameter k. Walker suggested using the DFFITS measure for the π-weights and he tried two variations of the ψ-function. A monotonic function (Huber's t) and a redescending function (Tukey's biweight) were compared.

METHODOLOGY

The discussion in this chapter will focus on the approach that will be used to answer the research objective. The two main elements of the objective are the development of a new combined estimator and a comparison of this estimator with competing combined estimators. Each element will be discussed in turn. Some of the answers are not presently known and will be determined in the process of the research effort. The plan of attack will be detailed here as best as possible. The literature review revealed some of the more promising estimators that will be candidates for the combined estimator and also possibilities for comparison techniques used in the simulation.

The primary question that must be addressed is: *What will be the original contribution*? A combined biased-robust estimation technique will be proposed that builds on the successes of previous research and modifies the way some of the important components are derived so that the resulting estimator is both improved and unique. The first phase of the effort will be to design a multi-stage robust estimator that is different from current approaches. The second phase will be to

transform the new robust estimator into a biased-robust estimator by applying a biased estimation technique that has not been used in this framework. The details of this approach are discussed below.

In order to best describe the methodology in terms of which questions need to be answered to accomplish the major research objective, a series of questions and answers follow. This format allows the reader to see the proposed sequence of issues that need to be studied and some of the proposed methods for dealing with the issues. The investigative questions will be used to guide the discussion.

I. How will the biased-robust estimator be developed?

- The proposed approach is to develop a multi-stage robust estimator and combine a biasing technique such as ridge or generalized ridge regression to create a biased-robust estimator. Recall that the objective of the multi-stage estimator is to combine the desirable properties of equivariance, high breakdown point, efficiency and bounded-influence in forming a robust estimate. A multi-stage approach using LMS or LTS as a first step estimator and bounded-influence for the second stage might be a good idea.

  A. What characteristics are required of the two classes of estimators (robust and biased) in order to take a robust estimator and a biased estimator and form a combined biased-robust estimator? For each class of estimator,

     1. What are the strengths and weaknesses associated with the available techniques?

        - Robust - The strengths and weakness of the robust estimators relative to the criteria mentioned above are displayed next to each technique. The techniques with the most shaded boxes will be the first to try.

        - Biased - The criteria for the biased estimators will include: computational ease, proper amount of shrinkage, resistance to alignment of orthogonal coefficients, and resistance to level of noise (error variance). Regarding computational ease, considerations will be made for whether the method of computation can be linked with the robust method and how the biasing parameter is calculated. Ridge and generalized ridge are the more proven techniques. Methods for determining the biasing parameter are numerous, but the Hoerl and Kennard (1976) iterative

technique for ridge and the Hoerl and Kennard (1970) technique for generalized ridge are most often used. Several techniques will be tested.

2. What are the properties most desirable in an estimator?

- Some properties are essential in order for the estimator to be able to handle all types of problems in these areas. Robust properties of efficiency, high breakdown point and bounded influence are probably the most important. The biased property of the proper amount of shrinkage is also important.

3. Which estimator is the best relative to the desirable properties?

- As was mentioned in the scope section, both statistical properties and performance are important. The estimators with the best statistical properties that can also be combined into biased-robust estimators will be selected for Monte Carlo simulation.

B. What characteristics are required for the biased-robust estimator?

1. What are the properties most desirable for the combined estimator?

- Obviously, the crucial element is that the combined estimator must be able to be computed. The next concern is whether the combined estimator works well in the simultaneous problem. In other words, if the robust portion is effective against outliers and the biased portion is effective against collinearity, will the combined method be effective in the presence of both problems?

2. Which estimator is the best relative to the desirable properties?

3. What are the challenges associated with combining the estimators?

- Computationally combining the techniques and maintaining the properties of each technique when they are put together is a challenge. For example, if a high breakdown point estimator is used in the first stage, will the combined estimator also have a high breakdown point?

II. Which estimators should be used for comparison in the performance test?

- The candidates will be chosen from two pools. The first pool will be those estimators that have performed the best in previous biased-robust studies. Examples include the ridge M-

estimate method of Askin and Montgomery (1984), the ridge LAV estimate of Pfaffenberger and Dielman (1990), and the ridge bounded-influence method of Walker (1987). The second pool of candidates will be the methods developed in this study.

III. How will each of these estimators be computed?

A. Is software available that generates some of the chosen estimators?

- In the robust case, software has been acquired or requested that computes various M-estimates and bounded-influence estimates. The software programs include SAS, LMSMVE (Dallal and Rousseeuw, 1992) and ROBSYS (Marazzi, 1987). The biased methods can be calculated using SAS.

B. Which estimators require coding?

- The combined estimator may require a programming language such as FORTRAN. SAS may also work.

C. Which programming language is most appropriate to code the remaining estimators?

- To be determined.

IV. How will the Monte Carlo simulation be developed to compare the estimators?

A. What characteristics of the data are important to vary in the simulation?

- The following characteristic were used by Askin and Montgomery (1984) and appear appropriate for this study. They include: type of error distribution, sample size, eigenvalue spread, alignment of the orthogonal coefficients, and noise level.

B. What type of design will be used in this experiment?

- Hopefully some sort of factorial design. The number of levels across factors probably will not be equal, so some sort of mixed level experiment will be appropriate.

C. How will the data be generated?

- Askin used FORTRAN to generate the data for their experiment, but he suggested a higher level language such as MathCAD or Mathematica.

V. What criteria will be used to measure the performance of the biased-robust estimators?

A. What performance indices are important?

- Of overall interest are the parameter estimates themselves relative to the true values. Another metric would be the prediction capability. In terms of dealing with outliers, one might be interested in identification of and dealing with highly influential points. Regarding the collinearity diagnostics, an index might be the level of variance reduction in the coefficients.

B. What measures can be calculated based on the simulation results?

- In Monte Carlo simulation the true values of the parameters estimates are known, so the analyst is very fortunate to be able to compare observed and actual values. Some of the measures used in previous research tests have been mean square error inefficiency ratios and mean absolute deviation ratios. Comparisons across techniques are also of interest, such as the number of times one technique estimates better than the other. If some of the data can be held back for prediction purposes, it may be interesting to calculate a PRESS statistic. Comparison of the level of downweighting of influence points may indicate the robust performance. There may also be a way to calculate the variance of the estimates based on distributional property assumptions.

Obviously, not all of the questions were fully answered. There is much to be learned during the research process. Hopefully, enough is now known so that the probability of contributing to the state of the art in regression is high. Two other possible extensions to the work presented here are: 1) applying these techniques to one or more real data sets and 2) applying the robust techniques to probability plotting for parameter estimation in reliability.

References

Adichie, J. N. (1967), "Estimation of Regression Coefficients Based on Rank Tests", *Annals of Mathematical Statistics*, 38, 894-904.

Akritas, M. G. (1991), "Robust M Estimation in the Two-Sample Problem", *Journal of the American Statistical Association*, 86, 201-204.

Andrews, D. F. (1974), "A Robust Method for Multiple Linear Regression", *Technometrics*, 16, 523-531.

Askin, R. G., and Montgomery, D. C. (1980), "Augmented Robust Estimators", *Technometrics*, 22, 333-341.

Askin, R. G., and Montgomery, D. C. (1984), "An Analysis of Constrained Robust Regression Estimators", *Naval Research Logistics Quarterly*, 31, 283-296.

Beckman, R. J., and Cook, R. D. (1983), "Outlier..........s," *Technometrics*, 25, 2, 119-149.

Beaton, A. E., and Tukey, J. W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data", *Technometrics*, 16, 147-185.

Bickel, P. J. (1973), "On Some Analogues to Linear Combination of Order Statistics in the Linear Model", *Annals of Statistics*, 1, 597-616.

Birch, J. B., and Agard, D. B. (1993), "Robust Inference in Regression: A Comparative Study", *Communications in Statistics - Simulation*, 22, 217-244.

Boscovich, R. J., (1757), "De letteraria expeditione per pontifician ditionem, et synopsis amplioris operis ac habertur plura eius ex exemplaria etiam sensroum impressa", *Bononiensi Scientiarum et Artium Instituto Atque Academia Commentarii*, 4, 353-396.

Box, G. E. P. (1953), "Non-normality and Tests on Variances", *Biometrika*, 40, 318-335.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, NY.

Coakley, C. W. and Hettmansperger, T. P. (1993), "A Bounded Influence, High Breakdown, Efficient Regression Estimator," *Journal of the American Statistical Association*, 88, 872-880.

Cook, R. D., and Weisberg, S. (1982), *Residuals and Influence in Regression*, Chapman and Hall, NY.

Cook, R. D., Hawkins, D. M., and Weisberg, S. (1992), "Comparison of Model Misspecification Diagnostics Using Residuals From Least Mean of Squares and Least Median of Squares Fits", *Journal of the American Statistical Association*, 87, 419-424.

Dallal, G. E., and Rousseeuw, P. J. (1992), "LMSMVE: A Program for Least Median of Squares Regression and Robust Distances", *Computers and Biomedical Research*, 25, 384-391.

Dallal, G. E., and Rousseeuw, P. J. (1992), "Documentation for the LMSMVE Program".

Daniel, C., and Wood, F. S. (1980), *Fitting Equations to Data*, Wiley, NY.

Denby, L., and Mallows, C. L. (1977), "Two Diagnostic Displays for Robust Regression Analysis", *Technometrics*, 1-13.

Dodge, Y. (1984), "Robust Estimation of Regression Coefficients by Minimizing a Convex Combination of Least Squares and Least Absolute Deviations", *Computational Statistics Quarterly*, 139-153.

Dodge, Y. (1987), *An Introduction to Statistical Data Analysis L1-norm Based, Statistical Data Analysis Based on the L1-norm and Related Methods*, Dodge, Y. (Ed.), 1-22.

Dollinger, M. B., and Staudte, R. G. (1991), "Influence Functions of Iteratively Reweighted Least Squares Estimators", *Journal of the American Statistical Association*, 86, 709-716.

Edgeworth, F. Y., (1887), "On Observations Relating to Several Quantities", *Hermathena*, 6, 279-285.

Ellis, S. P., and Morgenthaler, S. (1992), "Leverage and Breakdown in L Regression", *Journal of the American Statistical Association*, 87, 143-148.

Forsythe, A. B. (1972), "Robust Estimation of Straight Line Regression Coefficients by Minimizing p-th Power Deviations", *Technometrics*, 14, 159-166.

Gentlemen, W. M. (1965), "Robust Estimation of Multivariate Location by Minimizing p-th Power Transformations", Ph.D. dissertation, Princeton University.

Gibbons, D. G. (1981), "A Simulation Study of Some Ridge Estimators", *Journal of the American Statistical Association*, 76, 131-139.

Gibbons, D. I., McDonald, G. C., and Gunst, R. F. (1987), "The Complementary Use of Regression Diagnostics and Robust Estimators", *Naval Research Logistics Quaterly*, 34, 109-131.

Green, P. J. (1984), "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives", *Journal of the Royal Statistical Society B*, 46, 149-170.

Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation", *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F. R. (1975), "Beyond Location Parameters: Robust Concepts and Methods", *Bulletin of the International Statistical Institute*, 46, 375-382.

Hampel, F. R. (1985), "The Breakdown Points of the Mean Combined With Some Rejection Rules", *Technometrics*, 27, 95-107.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, NY.

Harter, H. L. (1974-1976), "The Method of Least Squares and Some Alternatives", Parts I-VI, *International Statistical Review*, 42, 147-174, 235-264, 282, 43, 1-44, 125-190, 273-278, 269-272, 44, 113-159.

Hawkins, D. W., Bradu, D., and Kass, G. V. (1984), "Location of Several Outliers in Multiple-Regression Data Using Elemental Sets," *Technometrics*, 26, 3, 197-208.

Heiler, S. (1981), "Robust Estimates in Linear Regression - A Simulation Approach", *Computational Statistics*, Buning, H., Naeve, P., and De Gruyter (Ed.), Berlin, 115-136.

Hill, R. W. (1977), "Robust Regression When There Are Outliers in the Carriers", Unpublished Ph.D. dissertation, Harvard University, Boston, MA.

Hill, R. W., and Holland, P. W. (1977), "Two Robust Alternatives to Least-Squares Regression", *Journal of the American Statistical Association*, 72, 828-833.

Hoaglin, D. C., and Welsch, R. E. (1978), "The Hat Matrix in Regression and ANOVA", *American Statistician*, 32, 17-22.

Hoerl, A. E., and Kennard, R. W. (1970a), "Ridge Regression: Biased Estimation for Non-orthogonal Problems", *Technometrics*, 12, 55-67.

Hoerl, A. E., and Kennard, R. W. (1970b), "Ridge Regression: Applications to Non-orthogonal Problems", *Technometrics*, 12, 69-82.

Hoerl, A. E. and Kennard, R. W. (1976), "Ridge Regression: Iterative Estimation of the Biasing Parameter", *Communications in Statistics*, A5, 77-88.

Hogg, R. V. (1974), "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory", *Journal of the American Statistical Association*, 69, 909-923.

Hogg, R. V. (1979), "An Introduction to Robust Estimation", *Robustness in Statistics*, Launer, R. L. and Wilkinson, G. N. (Ed.), 1-17.

Holland, P. W. (1973), "Weighted Ridge Regression: Combining Ridge and Robust Regression Methods", NBER Working Paper Series, Working Paper #11, 1-19.

Holland, P. W. (1977), "Robust Regression Using Iteratively Reweighted Least-Squares", *Communications in Statistics - Theory and Methods*, A6, 813-827.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter", *Annals of Mathematical Statistics*, 35, 73-101

Huber, P. J. (1972), "Robust Statistics: A Review", *Annals of Mathematical Statistics*, 43, 1041-1067.

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures, and Monte Carlo", *The Annals of Statistics*, 1, 799-821.

Huber, P. J. (1987), "The Place of the L1-norm in Robust Estimation", *Statistical Data Analysis Based on the L1-norm and Related Methods*, Yadolah Dodge (Ed.), 23-33.

Jaeckel, L. A. (1972), "Estimating Regression Coefficients by Minimizing the Dispersion of Residuals", *Annals of Mathematical Statistics*, 5, 1449-1458.

Jureckova, J. (1971), "Nonparametric Estimate of Regression Coefficients", *Annals of Mathematical Statistics*, 42, 1328-1338.

Koenker, R., and Bassett, G. J. (1978), "Regression Quantiles", *Econometrica*, 46, 33-50.

Krasker, W. S. (1980), "Estimation in Linear Regression Models with Disparate Data Points", *Econometrica*, 48, 1333-1346.

Krasker, W. S., and Welsh, R. E. (1982), "Efficient Bounded-Influence Regression Estimation", *Journal of the American Statistical Association*, 77, 595-603.

Launer, R. L., and Wilkinson, G. N. (Ed.) (1979), *Robustness in Statistics*, Academic Press, NY.

Lawrence, K. D., and Marsh, L. C. (1984), "Robust Ridge Estimation Methods for Predicting U. S. Coal Mining Fatalities", *Communications in Statistics-Theory and Methods*, 13, 139-149.

Lawrence, K. D., and Arthur, J. L. (Ed.) (1990), *Robust Regression: Analysis and Applications*, Marcel Dekker, NY.

Mallows, C. L. (1973), "Some Comments on Cp", *Technometrics*, 15, 661-675.

Mallows, C. L. (1975), "On Some Topics in Robustness", Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.

Marazzi, A. (1987), "Solving Bounded Influence Regression Problems with ROBSYS", *Statistical Data Analysis Based on the L1-norm and Related Methods*, Dodge, Y. (Ed.), 145-161.

Markatou, M., and Hettmansperger, T. P. (1990), "Robust Bounded-Influence Tests in Linear Models", *Journal of the American Statistical Association*, 85, 187-190.

Maronna R., Yohai, V. J. (1991), "The Breakdown Point of Simultaneous General M Estimates of Regression and Scale", *Journal of the American Statistical Association*, 86, 699-703.

McDonald, G. C., and Galarneau, D. I. (1975), "A Monte Carlo Evaluation of Some Ridge-type Estimators", *Journal of the American Statistical Association*, 70, 407-416.

Montgomery D. C., and Askin, R. G. (1981), "Problems of Nonnormality and Multicollinearity for Forecasting Methods Based on Least Squares", *AIIE Transactions*, 13, 102-115.

Montgomery, D. C., and Friedman, D. J. (1993), "Prediction Using Regression Models with Multicollinear Predictor Variables", *IIE Trransactions*, 25, 73-85.

Montgomery, D. C., and Peck, E. A. (1992), *Introduction to Linear Regression Analysis (2nd Ed)*, John Wiley and Sons, NY.

Pariente, S., and Welsch, R. E. (1977), "Ridge and Robust Regression Using Parametric Linear Programming", Working Paper, Alfred P. Sloan School of Management, MIT.

Pfaffenberger, R. C., and Dielman, T. E. (1985), "A Comparison of Robust Ridge Estimators", *Business Economics Section Proceedings of the American Statistical Association*, 631-635.

Pfaffenberger, R. C., and Dielman, T. E. (1990), "A Comparison of Regression Estimators When Both Multicollinearity and Outliers Are Present", *Robust Regression: Analysis and Applications*, Lawrence, K. D. and Arthur, J. L. (Ed.), 243-270.

Rey, W. J. J. (1983), *Introduction to Robust and Quasi-Robust Statistical Methods*, Springer-Verlag, Berlin, GE.

Ronchetti, E. (1987), "Bounded Influence Inference in Regression: A Review", *Statistical Data Analysis Based on the L1-norm and Related Methods*, Dodge, Y. (Ed.), 65-80.

Rousseeuw, P. J. (1984), "Least Median of Squares Regression", *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J., and Bassett, G. W., Jr. (1990), "The Remedian: A Robust Averaging Method for Large Data Sets", *Journal of the American Statistical Association*, 85, 97-104.

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, Wiley, NY.

Rousseeuw, P. J., and van Zomeren, B. C. (1990), "Unmasking Multivariate Outliers and Leverage Points", *Journal of the American Statistical Association*, 85, 633-651.

Ruppert, D., and Carroll, R. J. (1980), "Trimmed Least Squares Estimation in the Linear Model", *Journal of the American Statistical Association*, 75, 828-838.

SAS Institute (1990), *SAS/STAT User's Guide, Version 6, Fourth Ed., Volume 2*, SAS Institute, Cary, NC.

Simpson, D. G., Ruppert, D., and Carroll, R. J. (1992), "On One-Step GM Estimates and Stability of Influences in Linear Regression", *Journal of the American Statistical Association*, 87, 439-450.

Sposito, V. A., Kennedy, W. J., and Gentle, J. E. (1977), "Lp norm Fit of a Straight Line", *Applied Statitistics*, 26, 114-116.

Stigler, S. M. (1973), "Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920", *Journal of the American Statistical Association*, 68, 872-879.

Street, J. O., Carroll, R. J., and Ruppert, D. (1988), "A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares", *The American Statistician*, 42, 152-154.

Tukey, J. W. (1960), "A Survey of Sampling from Contaminated Distributions", *Contributions to Probability and Statistics*, Olkin (Ed.), Stanford University Press, Stanford, CA, 448-485.

Walker, E. (1987), "Augmented Bounded-Influence Estimators", *Statistical Computing Section Proceedings of the American Statistical Association*, 158-164.

Walker, E., and Birch, J. B. (1988), "Influence Measures in Ridge Regression", *Technometrics*, 30, 221-227.

Wegman, E. J., and Carroll, R. J. (1977), "A Monte Carlo Study of Robust Estimators of Location", *Communications in Statistics - Theory and Methods*, A6, 795-812.

Yale C., and Forsythe, A. B. (1976), "Winsorized Regression", *Technometrics*, 18, 291-300.

Yohai, V. J. (1987), "High Breakdown-Point and High Efficiency Robust Estimates for Regression", *The Annals of Statistics*, 15, 642-656.